

# Rényi Information Complexity and an Information Theoretic Characterization of the Partition Bound

Manoj M. Prabhakaran  
Department of Computer Science  
University of Illinois  
Urbana-Champaign, IL  
mmp@illinois.edu

Vinod M. Prabhakaran  
School of Technology and Computer Science  
Tata Institute of Fundamental Research  
Mumbai, India.  
vinodmp@tifr.res.in

June 29, 2016

## Abstract

In this work we introduce a new information-theoretic complexity measure for 2-party functions, called Rényi information complexity. It is a lower-bound on communication complexity, and has the two leading lower-bounds on communication complexity as its natural relaxations: (external) information complexity and logarithm of partition complexity. These two lower-bounds had so far appeared conceptually quite different from each other, but we show that they are both obtained from Rényi information complexity using two different, but natural relaxations:

1. The relaxation of Rényi information complexity that yields information complexity is to change the order of Rényi mutual information used in its definition from infinity to 1.
2. The relaxation that connects Rényi information complexity with partition complexity is to replace protocol transcripts used in the definition of Rényi information complexity with what we term “pseudo-transcripts,” which omits the interactive nature of a protocol, but only requires that the probability of any transcript given inputs  $x$  and  $y$  to the two parties, factorizes into two terms which depend on  $x$  and  $y$  separately. While this relaxation yields an apparently different definition than (log of) partition function, we show that the two are in fact identical. This gives us a surprising characterization of the partition bound in terms of an information-theoretic quantity.

We also show that if both the above relaxations are simultaneously applied to Rényi information complexity, we obtain a complexity measure that is lower-bounded by the (log of) relaxed partition complexity, a complexity measure introduced by Kerenidis et al. (FOCS 2012). We obtain a sharper connection between (external) information complexity and relaxed partition complexity than Kerenidis et al., using an arguably more direct proof.

Further understanding Rényi information complexity (of various orders) might have consequences for important direct-sum problems in communication complexity, as it lies between communication complexity and information complexity.

## 1 Introduction

Communication complexity, since the seminal work of Yao [26], has been a central question in theoretical computer science. Many of the recent advances in this area have centered around the notion of information complexity, which measures the *amount of information* about the inputs – rather than the *number of bits* – that should be present in a protocol’s transcript, if it should compute a function (somewhat) correctly. The more traditional approach for lower bounding communication complexity relied on *combinatorial complexity measures* of functions. The goal of this work is to relate these two lines of studying communication complexity with each other.

Currently, the two leading lower bounds for communication complexity in the literature come from these two lines: (external) information complexity  $IC$  [8, 2] and partition complexity  $\text{prt}$  [14]. Either of these two lower bounds upper-bounds (and hence gives an equally good or better lower bound than) all the other bounds used in the literature. An intriguing problem in this area has been to understand if one of these two bounds is a better lower-bound than the other. An important motivation behind this problem is the possibility of separating  $IC$

from communication complexity via an intermediate combinatorial lower bound, which will have consequences for direct-sum results in communication complexity (since  $IC$  is known to be equal to amortized communication complexity [5, 4]).

Kerenidis et al. [18] showed that information complexity “subsumes” (the logarithm of) a relaxed variant of partition complexity,  $\overline{\text{prt}}$ , in the sense that any lower bound on  $\log \overline{\text{prt}}$  in fact yields a lower bound on information complexity. Thus bounding  $\log \overline{\text{prt}}$  cannot yield stronger lower bounds than bounding information complexity. In turn, all the combinatorial bounds in the literature – other than  $\log \overline{\text{prt}}$  – are subsumed by  $\log \overline{\text{prt}}$ . On the other hand, in recent breakthrough results, Ganor, Kol and Raz [10, 11] showed that for a certain range of parameters, combinatorial lower bounds can be significantly stronger than information complexity lower bounds.<sup>1</sup> It remains open if such separations are possible for a less restrictive range of parameters (e.g., with communication complexity that is say, super-logarithmic in the input size). In the absence of a result analogous to that of [18] for  $\text{prt}$  itself,  $\text{prt}$  remains a candidate for showing such separations.

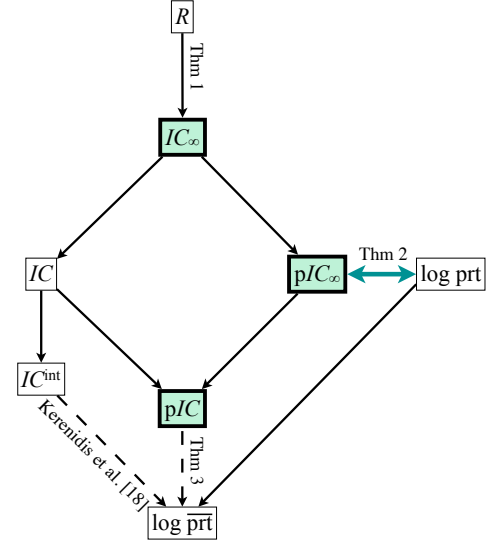
In this work, we do not pursue the question of whether  $\log \text{prt}$  could be larger than  $IC$  or vice versa. Instead, we develop a new information-theoretic complexity measure,  $IC_\infty$  which is as large or larger than both  $IC$  and  $\log \text{prt}$  (see Figure 1), and has *natural* relaxations that yield  $IC_\infty$  and  $\log \text{prt}$  respectively.  $IC_\infty$  is thus a candidate for separating  $IC$  and communication complexity for a larger range of parameters than currently known to be possible. Further, the relaxation of  $IC_\infty$  to  $\log \text{prt}$  reveals a surprising information-theoretic definition for  $\text{prt}$ . Since this new definition of ( $\log$  of)  $\text{prt}$  has a markedly different form, we give it a different name,  $\text{p}IC_\infty$ .

We also consider applying *both the relaxations* mentioned above simultaneously to  $IC_\infty$ . This yields a new complexity measure  $\text{p}IC$ . We then show that  $\text{p}IC$  is essentially lower bounded by  $\log \overline{\text{prt}}$ , the relaxed partition complexity. *This recovers a result similar to that of [18], but with sharper parameters and an arguably more direct proof.*<sup>2</sup>

The relation between the new and old complexity measures are shown in Figure 1. (Also see Figure 3 for further extensions.) The new complexity measures are informally described below.

**Rényi Information Complexity.** (External) Information complexity of a function is defined as the mutual information between the transcript and the inputs, and is a lower bound on the communication complexity of the function. The notion of mutual information in this definition is due to Shannon. Rényi mutual information  $I_\alpha(A; B)$ , parametrized by  $\alpha \geq 0$ , is a generalization of Shannon’s mutual information (see [25] for a recent treatment), with the latter corresponding to  $\alpha \rightarrow 1$ . We observe that information complexity continues to be a lower bound on communication complexity for all values of  $\alpha$ . In particular, we may consider  $I_\infty$  instead of  $I_1$  to define information complexity. The resulting notion of information complexity will be called  $IC_\infty$ .

**Pseudotranscript Complexity.** Communication complexity, as well as information complexity, is defined in terms of a protocol. In contrast, the more traditional combinatorial lower bounds on communication complexity are defined in terms of simpler combinatorial properties of the function’s truth table. We propose complexity measures based on one such property (which has been widely used in the analysis of protocols, but to the best of our knowledge, has never been isolated to define a complexity measure of functions).



**Figure 1** New complexity measures (shaded) and their relation to existing ones. Existing ones shown include the (public-coin) worst-case communication complexity ( $R$ ), external and internal information complexity ( $IC$  and  $IC^{\text{int}}$ ), partition complexity ( $\text{prt}$ ) and relaxed partition complexity ( $\overline{\text{prt}}$ ). An arrow from one measure to another shows that the latter is a lower-bound for the former. (The dashed lines indicate that the lower bound holds up to constant factors and shifts in error bounds.)  $\text{p}IC_\infty$  is exactly equal to  $\log \overline{\text{prt}}$ .

<sup>1</sup>These results use combinatorial lower bounds to establish that communication complexity could be exponentially larger than information complexity. The communication complexity in these examples is (sub-)logarithmic in the size of the input itself.

<sup>2</sup>Our result does not subsume the result of Kerenidis et al. [18], as they deal with internal information complexity, while it is more natural for us to work with external information complexity. Conversely, the result of [18] does not yield our result for external information complexity (due to the parameters), nor the relation with the intermediate complexity measure  $\text{p}IC$ .

Consider a function (generalized later to relations)  $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ . We define a random variable  $Q$  over a space  $\mathcal{Q}$  to be a *pseudotranscript* for  $f$  if there exist two functions  $\alpha : \mathcal{Q} \times \mathcal{X} \rightarrow \mathbb{R}^+$  and  $\beta : \mathcal{Q} \times \mathcal{Y} \rightarrow \mathbb{R}^+$ , such that  $\Pr[Q = q | X = x, Y = y] = \alpha(q, x)\beta(q, y)$ , for all  $q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}$ . This definition is motivated by the fact that the transcripts in a protocol do satisfy it (see [Footnote 7](#)). However, a pseudotranscript need not correspond to a protocol (indeed, any “tiling” of a function’s table yields a pseudotranscript, but it need not correspond to a valid protocol). We also associate a value  $z_q$  with a pseudotranscript  $q$ ; the error  $\text{err}_{f,Q}$  is defined in terms of the probability of this value matching the function’s output. We do not include any other properties of a protocol in defining a pseudotranscript.

We can define complexity measures  $\text{pIC}$  and  $\text{pIC}_\infty$  as relaxations of  $\text{IC}$  and  $\text{IC}_\infty$ , simply by replacing protocols in their definitions with pseudotranscripts.

**Relations Among the Complexities.** The main results in this work, apart from introducing the new complexity measures, are connections between  $\text{pIC}_\infty$  and  $\text{prt}$  and between  $\text{pIC}$  and  $\overline{\text{prt}}$ .

- Firstly, we show that  $\text{pIC}_\infty = \log \text{prt}$ .  $\text{pIC}_\infty$  and  $\text{prt}$  are defined very differently.  $\text{prt}$  is concerned with *tiling* the function table with weighted tiles: a tile  $t$  is a rectangle in the input domain along with an output value  $z_t$ .  $\text{prt}$  is the minimum total weight of tiles needed such that for each input  $(x, y)$ , the weight of the tiles covering it adds up to 1, and the weight of the tiles with  $z_t \neq f(x, y)$  is below the error threshold  $\mathcal{E}(x, y)$ .<sup>3</sup> On the other hand,  $\text{pIC}_\infty$  relates to pseudotranscripts  $q$ , which are similar to tiles in that they define a value  $z_q$  and a rectangle of all  $(x, y)$  such that  $p(q|x, y) > 0$ , but are more general in that there is no single “weight” on such a rectangle. Given our definitions, it is not hard to see that  $\log \text{prt}$  is as large or larger than  $\text{pIC}_\infty$ , as any tiling can be naturally interpreted as a pseudotranscript  $Q$  with the same error, and in that case, the log of the value of the tiling indeed equals  $I_\infty(X, Y; Q)$ . What is more surprising is that any pseudotranscript  $Q$  can be converted to a tiling of the appropriate value (and same error). This conversion “slices” an uneven weight function  $p(q|x, y)$  over a rectangle into weights  $\omega_{q,t}$  over tiles  $t$  inside the rectangle; the weight of a tile  $t$  is the sum of the contributions to its weight from all the different values of  $q$ :  $w(t) = \sum_q \omega_{q,t}$ . Then it turns out that the value of the tiling so obtained will be equal to  $I_\infty(X, Y; Q)$ .

This equivalence gives a new perspective on the partition complexity. Firstly, it shows that partition complexity exploits *exactly* the properties of a pseudotranscript, which is not apparent from its original definition. Secondly, it gives an information theoretic interpretation of a complexity measure defined in a traditional combinatorial manner. This is the first instance of the two lines of lower-bounding techniques for communication complexity – information theoretic and combinatorial – converging.

- Our second main result is that lower bounds on  $\log \overline{\text{prt}}$  are in fact lower bounds on  $\text{pIC}$ . More precisely, we show that  $\text{pIC}(f, \varepsilon) \geq \delta \log \overline{\text{prt}}(f, \varepsilon + \delta) - (\delta \log \log |\mathcal{X}||\mathcal{Y}| + 3)$ . This is along the same lines as the result of [18], with improved parameters (in [18], the multiplicative overhead in the leading term is  $\delta^2$  instead of  $\delta$ ).

The proof of this result is technically more involved, but is closely based on the simple slicing construction from the above result. The high-level idea is to first slice  $p(q|x, y)$  into weights  $\omega_{q,t}$  for each tile  $t$ , and then discard the contributions to  $w(t)$  from those  $\omega_{q,t}$  which are too large. One needs to ensure that the weight of the tiles discarded in this fashion is small (as it contributes to the error), while the weight of the remaining tiles is also small (as it contributes to the value of the tiling). For the first part, we show how (Shannon’s) mutual information  $I(X, Y; Q)$  can be approximated by a convex combination of non-negative values, and then apply Markov’s inequality. For the second part, we rely on a geometric argument to derive a bound on the weight of the remaining tiles.

## 1.1 Related Work

Many of the recent advances in the field of communication complexity [26] have followed from using various notions of information complexity. Earlier notions of information complexity appeared implicitly in several works [1, 20, 23], and was first explicitly defined in [8] and further developed in [2]. Information complexity has been extensively used or studied in the recent communication complexity literature (e.g., [5, 4, 6, 7, 18, 3, 10, 9, 11]). The notion was also adapted to specialized models or tasks [17, 15, 16, 12].

The partition bound was developed in [14], and has subsumed a long line of combinatorial bounds [19] (see e.g., [14, 9]). The relaxed partition bound put forth in [18], similarly subsumes several combinatorial bounds, with the exception of the partition bound itself.

---

<sup>3</sup>For  $\text{prt}$ , as well as  $\text{pIC}_\infty$  and  $\text{IC}_\infty$ , we use a very general notion of error, in which the error is specified as a function  $\mathcal{E} : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ .

In 1960, generalizing Shannon’s entropy, Rényi proposed new measures of entropy and divergence [22], now known after him. Subsequently, several authors developed different notions of mutual information based on these measures. One such definition attributed to Sibson [24] has recently come to be regarded as the most standard choice [25], and this is the basis for our definition of  $I_\infty(A : B)$ . Properties of  $I_\alpha$  for various parameters  $\alpha$  have been studied in [13, 25]. In information theory literature, the use of generalized notions of mutual information to obtain strong lower bounds for “one-shot” versions of communication problems (rather than amortized/direct-sum versions where Shannon’s mutual information is often appropriate) has a long history starting with the work of Ziv and Zakai [28, 27]. In the communication complexity literature, Rényi divergence was used as a technical tool in deriving one of the results in [2].

Recently, the authors of this work proposed a distributional complexity measure, *Wyner tension* (or more generally, *tension gap*) which is a lower bound for information complexity [21]. We leave it for future work to explore the exact connections between these bounds and the ones in the current work. We mention that for the case when the inputs are independent, Wyner tension is identical to  $pIC^{\text{int}}$  (defined in Section 6), and a result in [21] is subsumed by the results in this work.

## 2 Preliminaries

Let  $f : \mathcal{X} \times \mathcal{Y} \rightarrow 2^{\mathcal{Z}}$  be a relation. Alice who has input  $x \in \mathcal{X}$  and Bob who has input  $y \in \mathcal{Y}$  want to output any  $z \in f(x, y)$ . We consider public-coin protocols, in which Alice and Bob have access to a common random string independent of the inputs; they may also use private local randomness. For such a protocol  $\pi$ , we say that the probability of error, which we view as a **function** of  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ , is

$$\text{err}_{f,\pi}(x, y) = \Pr[\pi(x, y) \notin f(x, y)],$$

where  $\pi(x, y)$  is the *output* of the protocol and the probability is over the randomness in the protocol execution.<sup>4</sup> An error function  $\mathcal{E}$  that is of particular interest is the constant (or worst-case) error function:  $\mathcal{E}(x, y) = \varepsilon$  for some constant  $\varepsilon$ , for all  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ .

For a protocol  $\pi$ , let  $\#\text{bits}(\pi, x, y)$  denote the maximum number of bits exchanged in an execution of  $\pi$  with inputs  $(x, y)$ , in the worst case (i.e., over all choices of randomness). Note that this measure excludes the number of bits in the public randomness. The (worst case) *communication complexity*  $R(f, \mathcal{E})$  of  $f$ , for an error function  $\mathcal{E}$ , is defined as

$$R(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f,\pi} \leq \mathcal{E}}} \max_{x,y} \#\text{bits}(\pi, x, y).$$

To define information complexities, we will need to consider the distribution  $\mathbf{p}_{X,Y}$  on the inputs  $X, Y$ . Let  $\Pi$  be the random variable that denotes the communication transcript *and* the public-coins of the protocol  $\pi$ . Then, the external information cost of the protocol  $\pi$  under the input distribution  $\mathbf{p}_{X,Y}$  is  $I(X, Y; \Pi)$ , i.e., the amount of information about the inputs  $X, Y$  contained in  $\Pi$ . The (non-distributional) *external information complexity*  $IC(f, \mathcal{E})$  is defined as

$$IC(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f,\pi} \leq \mathcal{E}}} \max_{\mathbf{p}_{X,Y}} I(X, Y; \Pi).$$

Similarly, internal information complexity is defined as

$$IC^{\text{int}}(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f,\pi} \leq \mathcal{E}}} \max_{\mathbf{p}_{X,Y}} I(X; \Pi|Y) + I(Y; \Pi|X).$$

Here the internal information cost,  $I(X; \Pi|Y) + I(Y; \Pi|X)$ , of the protocol  $\pi$  under input distribution  $\mathbf{p}_{X,Y}$  is the sum of the information learned by the parties about each other’s input from  $\Pi$ . The following relationship between these quantities is well-known.

$$IC^{\text{int}}(f, \mathcal{E}) \leq IC(f, \mathcal{E}) \leq R(f, \mathcal{E}).$$

A *tile* for  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$  is a pair  $(r_X \times r_Y, z)$ , where  $r_X \subseteq \mathcal{X}$ ,  $r_Y \subseteq \mathcal{Y}$  and  $z \in \mathcal{Z}$ . If  $t = (r_X \times r_Y, z)$ , then we let  $\mathcal{X}_t, \mathcal{Y}_t$ , and  $z_t$  denote  $r_X, r_Y$  and  $z$  respectively. We say  $(x, y) \in t$  if and only if  $x \in \mathcal{X}_t$  and  $y \in \mathcal{Y}_t$ . The set of all tiles for  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$  is denoted by  $\mathcal{T}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$  or simply  $\mathcal{T}$  (if  $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$  are clear from the context).

<sup>4</sup>For a protocol to be considered valid, we will insist that the two parties output the same value with probability 1; hence the output of a protocol is well-defined.

For a relation  $f : \mathcal{X} \times \mathcal{Y} \rightarrow 2^{\mathcal{Z}}$  and probability of error  $\mathcal{E} : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ , the partition complexity [14] is defined as follows:<sup>5</sup>

$$\begin{aligned} \text{prt}(f, \mathcal{E}) = \min_{w: \mathcal{T} \rightarrow [0, 1]} \sum_{t \in \mathcal{T}} w(t) \quad \text{subject to} \\ \sum_{t \in \mathcal{T}: (x, y) \in t} w(t) = 1, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \end{aligned} \quad (1)$$

$$\sum_{\substack{t \in \mathcal{T}: (x, y) \in t, \\ z_t \in f(x, y)}} w(t) \geq 1 - \mathcal{E}(x, y), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (2)$$

For a weight function  $w$  as above, we define the error function as  $\text{err}_{f, w}(x, y) = \sum_{\substack{t \in \mathcal{T}: (x, y) \in t, \\ z_t \notin f(x, y)}} w(t)$ ; then the condition (2) can be written as a condition on this error function:  $\text{err}_{f, w} \leq \mathcal{E}$ .

The relaxed partition complexity [18] relaxes the equality constraint in (1) to an inequality. Further, the error function is restricted to be a constant function given by  $\mathcal{E}(x, y) = \varepsilon$ . Specifically, for a relation  $f$  and a constant  $0 \leq \varepsilon \leq 1$ ,

$$\begin{aligned} \overline{\text{prt}}(f, \varepsilon) = \min_{w: \mathcal{T} \rightarrow [0, 1]} \sum_{t \in \mathcal{T}} w(t) \quad \text{subject to} \\ \sum_{t \in \mathcal{T}: (x, y) \in t} w(t) \leq 1, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \end{aligned} \quad (3)$$

$$\sum_{\substack{t \in \mathcal{T}: (x, y) \in t, \\ z_t \in f(x, y)}} w(t) \geq 1 - \varepsilon, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (4)$$

The distributional form of relaxed partition complexity is defined for a distribution  $\mu$  and  $\varepsilon \in [0, 1]$  as follows:

$$\begin{aligned} \overline{\text{prt}}^\mu(f, \varepsilon) = \min_{w: \mathcal{T} \rightarrow [0, 1]} \sum_{t \in \mathcal{T}} w(t) \quad \text{subject to} \\ \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \quad \sum_{t \in \mathcal{T}: (x, y) \in t} w(t) \leq 1, \\ \sum_{x, y} \mu(x, y) \sum_{\substack{t \in \mathcal{T}: (x, y) \in t, \\ z_t \in f(x, y)}} w(t) \geq 1 - \varepsilon. \end{aligned}$$

For a weight function  $w$  as above and a distribution  $\mu$  over  $\mathcal{X} \times \mathcal{Y}$ , we write  $\overline{\text{err}}_{f, w}^\mu$  for  $1 - \sum_{x, y} \mu(x, y) \sum_{\substack{t \in \mathcal{T}: (x, y) \in t, \\ z_t \in f(x, y)}} w(t)$ ; so the second condition can be written as  $\overline{\text{err}}_{f, w}^\mu \leq \varepsilon$ . As shown in [18],  $\overline{\text{prt}}(f, \varepsilon) = \max_\mu \overline{\text{prt}}^\mu(f, \varepsilon)$ .

### 3 Rényi Information Complexity and Pseudotranscripts

In this section we define our new complexity measures.

**Rényi information complexity.** For a pair of random variables  $(A, B)$  over  $\mathcal{A} \times \mathcal{B}$ , Rényi mutual information of order  $\infty$  is defined as (see, e.g., [25])

$$I_\infty(A; B) = \log \left( \sum_{b \in \mathcal{B}} \max_{a \in \mathcal{A}: \mathbf{p}_A(a) > 0} \mathbf{p}_{B|A}(b|a) \right).$$

For a protocol  $\pi$  and an input distribution  $\mathbf{p}_{X, Y}$ , we will call  $I_\infty(X, Y; \Pi)$  the Rényi information cost. *Rényi information complexity*  $IC_\infty(f, \mathcal{E})$  is defined as the smallest worst-case (over input distributions) Rényi information cost of any protocol which has a probability of error at most  $\mathcal{E}(x, y)$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ .

$$IC_\infty(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f, \pi} \leq \mathcal{E}}} \max_{\mathbf{p}_{X, Y}} I_\infty(X, Y; \Pi).$$

<sup>5</sup>The definition presented in [14] is slightly more restrictive in the kind of relations and error functions considered.

Note the above definition is identical to the definition of  $IC(f, \mathcal{E})$  except that  $I_\infty$  is used in place of mutual information  $I$ . It is easy to see that the inner maximization above is obtained by any input distribution  $\mathbf{p}_{X,Y}$  with full support. Hence, we may equivalently write

$$IC_\infty(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f, \pi} \leq \mathcal{E}}} I_\infty(X, Y : \Pi),$$

where we define  $I_\infty(A : B)$  which is a function only of  $\mathbf{p}_{B|A}$  as

$$I_\infty(A : B) = \log \left( \sum_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} \mathbf{p}_{B|A}(b|a) \right).$$

**Theorem 1.**  $IC(f, \mathcal{E}) \leq IC_\infty(f, \mathcal{E}) \leq R(f, \mathcal{E})$ .

*Proof.* The inequality  $IC(f, \mathcal{E}) \leq IC_\infty(f, \mathcal{E})$  follows from  $I(X, Y; \Pi) \leq I_\infty(X, Y; \Pi)$ , which in turn follows from the monotonicity of  $\alpha$ -mutual information [13, Theorem 4(b)]; for completeness, we give a proof that  $I(A; B) \leq I_\infty(A; B)$  in the [Appendix A.1](#).

The proof of  $IC_\infty(f, \mathcal{E}) \leq R(f, \mathcal{E})$  is simple. Consider any public-coin protocol  $\pi$ . Let  $\Pi = (\Phi, \Psi)$  where  $\Phi$  represents the public-coins and  $\Psi$  the transcript of  $\pi$ . W.l.o.g.,  $\Psi$  can be considered to be a deterministic function of  $\Phi$  and the inputs  $X, Y$ .<sup>6</sup> We write  $\Psi(x, y; \phi)$  to denote the transcript of  $\pi$  on inputs  $(x, y)$  and public coins  $\phi$ . Note that  $\#bits(\pi, x, y) = \max_\phi |\Psi(x, y; \phi)|$  (where  $|\cdot|$  denotes the length of a bit string). We shall show that  $I_\infty(X, Y : \Pi) \leq \max_{x, y, \phi} |\Psi(x, y; \phi)|$ . This suffices since

$$IC_\infty(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f, \pi} \leq \mathcal{E}}} I_\infty(X, Y : \Pi). \quad R(f, \mathcal{E}) = \inf_{\substack{\text{protocol } \pi: \\ \text{err}_{f, \pi} \leq \mathcal{E}}} \max_{x, y, \phi} |\Psi(x, y; \phi)|.$$

Note that  $\mathbf{p}_{\Phi\Psi|XY}(\phi, \psi|x, y) = \mathbf{p}_\Phi(\phi)\mathbf{p}_{\Psi|\Phi XY}(\psi|\phi, x, y)$ . Then,

$$\begin{aligned} I_\infty(X, Y : \Phi, \Psi) &= \log \sum_{\phi} \mathbf{p}_\Phi(\phi) \sum_{\psi} \max_{x, y} \mathbf{p}_{\Psi|\Phi XY}(\psi|\phi, x, y) \\ &\leq \log \max_{\phi} \sum_{\psi} \max_{x, y} \mathbf{p}_{\Psi|\Phi XY}(\psi|\phi, x, y) \\ &= \max_{\phi} \log |\{\psi : \exists (x, y) \text{ s.t. } \psi = \Psi(x, y; \phi)\}| \leq \max_{x, y, \phi} |\Psi(x, y; \phi)|. \end{aligned} \quad \square$$

**Pseudotranscript and pseudo-information complexities.** A random variable  $Q$  defined on an alphabet  $\mathcal{Q}$  and jointly distributed with the inputs  $X, Y$  is said to be a *pseudotranscript* if  $\mathbf{p}_{Q|X, Y}$  satisfies the following *factorization condition*:

$$\mathbf{p}_{Q|X, Y}(q|x, y) = \alpha(q, x)\beta(q, y), \quad \forall q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y},$$

for some pair of functions  $\alpha : \mathcal{Q} \times \mathcal{X} \rightarrow \mathbb{R}^+$  and  $\beta : \mathcal{Q} \times \mathcal{Y} \rightarrow \mathbb{R}^+$ . In addition, we will require that  $Q$  defines an output, i.e., for each  $q$  there is an associated  $z_q \in \mathcal{Z}$ .

For any protocol  $\pi$ , clearly,  $\Pi$ , which is composed of the public-coins and the transcript, is a pseudotranscript.<sup>7</sup> For a pseudotranscript  $Q$ , the probability of error is defined analogously to that for a protocol as

$$\text{err}_{f, Q}(x, y) = \Pr[z_Q \notin f(x, y) | (X, Y) = (x, y)].$$

<sup>6</sup>Any protocol using private randomness can be transformed to one with only public randomness, by including the private coins as part of the public-coins, without changing the number of bits communicated. Further, this can only increase the quantity  $I_\infty(X, Y; \Pi)$ . Hence, it is enough to prove the inequality after carrying out this transformation.

<sup>7</sup> $Q = \Pi$  satisfies the factorization condition, as in that case, for  $q = (\phi, m_1, \dots, m_t)$ ,  $\Pr[q|x, y] = \alpha(q, x) \cdot \beta(q, y)$ , where say,  $\alpha(q, x) = \Pr[\phi] \cdot \Pi_{\text{odd } i} \Pr[m_i | \phi, m_1, \dots, m_{i-1}, x]$ , and  $\beta(q, y) = \Pi_{\text{even } i} \Pr[m_i | \phi, m_1, \dots, m_{i-1}, y]$ . Also, we can associate the output of the protocol, which we insisted must be the same for both parties for a valid protocol, as the corresponding output  $z_Q$ . Though the output of the parties could in principle depend on the local input and local randomness, the factorization condition and the requirement that the outputs agree together imply that the output can be unambiguously determined from the transcript together with the public-coins.



We define the following “pseudo-quantities” corresponding to  $IC_\infty$  and  $IC$  where  $\Pi$  is replaced by pseudotranscripts:

$$\begin{aligned} pIC_\infty(f, \mathcal{E}) &= \inf_{\substack{\text{pseudotranscript } Q: \mathbf{p}_{X,Y} \\ \text{err}_{f,Q} \leq \mathcal{E}}} \max_{\mathbf{p}_{X,Y}} I_\infty(X, Y; Q) = \inf_{\substack{\text{pseudotranscript } Q: \\ \text{err}_{f,Q} \leq \mathcal{E}}} I_\infty(X, Y : Q) \\ pIC(f, \mathcal{E}) &= \inf_{\substack{\text{pseudotranscript } Q: \mathbf{p}_{X,Y} \\ \text{err}_{f,Q} \leq \mathcal{E}}} \max_{\mathbf{p}_{X,Y}} I(X, Y; Q). \end{aligned}$$

**Observation 1.** *Since, for any protocol, its transcript is a pseudotranscript as well, we have  $pIC_\infty(f, \mathcal{E}) \leq IC_\infty(f, \mathcal{E})$  and  $pIC(f, \mathcal{E}) \leq IC(f, \mathcal{E})$ . Furthermore, since  $I(A; B) \leq I_\infty(A; B)$ , we also have  $pIC(f, \mathcal{E}) \leq pIC_\infty(f, \mathcal{E})$ .*

## 4 $pIC_\infty$ Equals the Partition Bound

**Theorem 2.** *For any relation  $f : \mathcal{X} \times \mathcal{Y} \rightarrow 2^{\mathcal{Z}}$  and error function  $\mathcal{E}$ ,  $pIC_\infty(f, \mathcal{E}) = \log \text{prt}(f, \mathcal{E})$ .*

We prove  $pIC_\infty(f, \mathcal{E}) \leq \log \text{prt}(f, \mathcal{E})$  and  $pIC_\infty(f, \mathcal{E}) \geq \log \text{prt}(f, \mathcal{E})$  separately. The first direction is easy, and follows by considering the tiles in a given partition as the pseudo transcripts.

**Lemma 1.**  $pIC_\infty(f, \mathcal{E}) \leq \log \text{prt}(f, \mathcal{E})$ .

The proof of this lemma is given in [Appendix A.2](#). Now we turn to the other direction.

**Lemma 2.**  $pIC_\infty(f, \mathcal{E}) \geq \log \text{prt}(f, \mathcal{E})$ .

The proof of this lemma will also serve as a starting point in proving the result in [Section 5](#).

*Proof.* Suppose  $\mathbf{p}_{Q|X,Y}$  satisfies the factorization and output consistency conditions,  $\text{err}_{f,Q} \leq \mathcal{E}$  and  $pIC_\infty(f, \mathcal{E}) = I_\infty(X, Y : Q)$ . Let  $\mathcal{T}$  be the set of all tiles. To define the partition  $w : \mathcal{T} \rightarrow [0, 1]$ , we shall (in (8)) define quantities  $\omega_{q,t}$  (for  $(q, t) \in \mathcal{Q} \times \mathcal{T}$ ) and probability distribution  $\mathbf{p}_{T|Q,X,Y}$ , where  $T$  is a random variable over  $\mathcal{T}$ , such that the following conditions hold.

$$\omega_{q,t} = 0 \quad \forall (q, t) \in \mathcal{Q} \times \mathcal{T} \text{ s.t. } z_t \neq z_q \quad (5)$$

$$p(q, t|x, y) = \begin{cases} \omega_{q,t} & \text{if } (x, y) \in t \\ 0 & \text{otherwise} \end{cases} \quad \forall (q, t) \in \mathcal{Q} \times \mathcal{T}, (x, y) \in \mathcal{X} \times \mathcal{Y} \quad (6)$$

$$\log \sum_{q \in \mathcal{Q}, t \in \mathcal{T}} \omega_{q,t} = I_\infty(X, Y : Q) \quad (7)$$

Now, if we let  $w : \mathcal{T} \rightarrow [0, 1]$  be defined by  $w(t) = \sum_{q \in \mathcal{Q}} \omega_{q,t}$ , then it is easy to verify that (1) and (2) hold, and further  $\log \text{prt}(f, \mathcal{E}) \leq \log \sum_{t \in \mathcal{T}} w(t) = I_\infty(X, Y : Q) = pIC_\infty(f, \mathcal{E})$ .

Thus, to complete the proof, it suffices to define  $\mathbf{p}_{T|Q,X,Y}$  and  $\omega_{q,t}$  so that the above conditions (5)-(7) are satisfied. Recall that, since  $Q$  is a pseudotranscript,  $\mathbf{p}_{Q|X,Y}$  satisfies the factorization condition; i.e., we can write

$$\mathbf{p}_{Q|X,Y}(q|x, y) = \alpha(q, x)\beta(q, y), \quad \forall q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y},$$

for some pair of functions  $\alpha : \mathcal{Q} \times \mathcal{X} \rightarrow \mathbb{R}^+$  and  $\beta : \mathcal{Q} \times \mathcal{Y} \rightarrow \mathbb{R}^+$ . For  $q \in \mathcal{Q}$  and  $t \in \mathcal{T}$ , let

$$\sigma_{q,t} = \min_{x \in \mathcal{X}_t} \alpha(q, x) - \max_{x' \notin \mathcal{X}_t} \alpha(q, x') \quad \text{and} \quad \tau_{q,t} = \min_{y \in \mathcal{Y}_t} \beta(q, y) - \max_{y' \notin \mathcal{Y}_t} \beta(q, y').$$

Above, in defining  $\max_{x' \notin \mathcal{X}_t}$ , if no such  $x'$  exists – i.e.,  $\mathcal{X}_t = \mathcal{X}$  – we take the maximum to be 0 (and similarly for  $\max_{y' \notin \mathcal{Y}_t}$ ). Now, let

$$\begin{aligned} \mathcal{T}_q &= \{t \in \mathcal{T} \mid \sigma_{q,t} > 0, \tau_{q,t} > 0, \text{ and } z_q = z_t\} \\ \omega_{q,t} &= \begin{cases} \sigma_{q,t} \cdot \tau_{q,t} & \text{if } t \in \mathcal{T}_q \\ 0 & \text{if } t \notin \mathcal{T}_q. \end{cases} \\ p(t|x, y, q) &= \begin{cases} \sigma_{q,t} \cdot \tau_{q,t} \cdot \frac{1}{p(q|x, y)} & \text{if } (x, y) \in t, t \in \mathcal{T}_q \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (8)$$

We shall use the following claim (proven below) to show that  $\mathbf{p}_{T|X,Y,Q}$  is a valid probability distribution and that conditions (5)-(7) are satisfied.

**Claim 1.** For any  $q \in \mathcal{Q}$  and  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ ,  $\sum_{t \in \mathcal{T}_q: (x, y) \in t} \sigma_{q,t} \cdot \tau_{q,t} = p(q|x, y)$ .

To see that  $\mathbf{p}_{T|X,Y,Q}$  is a valid probability distribution, firstly we note that the quantity  $p(t|x, y, q)$  in (8) is well-defined: if  $t \in \mathcal{T}_q$  and  $(x, y) \in t$ , then  $\sigma_{q,t} > 0, \tau_{q,t} > 0$  and hence,  $p(q|x, y) = \alpha(q, x)\beta(q, y) > 0$ . Secondly, from Claim 1 it follows that  $\sum_{t \in \mathcal{T}} p(t|x, y, q) = 1$ .

Next, we verify the conditions (5)-(7). (5) directly follows from the definition of  $\omega_{q,t}$ . To see (6), we note that

$$p(q, t|x, y) = p(q|x, y) \cdot p(t|q, x, y) = \begin{cases} \sigma_{q,t} \cdot \tau_{q,t} & \text{if } (x, y) \in t, t \in \mathcal{T}_q \\ 0 & \text{if } (x, y) \in t, t \notin \mathcal{T}_q \\ 0 & \text{if } (x, y) \notin t \end{cases} = \begin{cases} \omega_{q,t} & \text{if } (x, y) \in t \\ 0 & \text{if } (x, y) \notin t \end{cases}$$

To see that (7) holds, fix a  $q \in \mathcal{Q}$ . Note that any  $t \in \mathcal{T}_q$ , if  $\sigma_{q,t} \cdot \tau_{q,t} > 0$ , then from the definition of  $\sigma_{q,t}$  and  $\tau_{q,t}$  it follows that  $(x^*, y^*) \in t$ , where  $x^* = \arg \max_{x \in \mathcal{X}} \alpha(q, x)$  and  $y^* = \arg \max_{y \in \mathcal{Y}} \beta(q, y)$ . Hence

$$\sum_{t \in \mathcal{T}} \omega_{q,t} = \sum_{t \in \mathcal{T}_q: (x^*, y^*) \in t} \sigma_{q,t} \cdot \tau_{q,t} = p(q|x^*, y^*),$$

where the last equality follows from Claim 1. But,  $p(q|x^*, y^*) = \max_{x \in \mathcal{X}, y \in \mathcal{Y}} \alpha(q, x)\beta(q, y) = \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p(q|x, y)$ . Thus,

$$\log \sum_{q \in \mathcal{Q}, t \in \mathcal{T}} \omega_{q,t} = \log \sum_{q \in \mathcal{Q}} \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p(q|x, y) = I_\infty(X, Y : Q).$$

□

To complete the proof of Lemma 2, we prove Claim 1.

*Proof of Claim 1.* Fix  $q \in \mathcal{Q}$ . Let  $\mathcal{X} = \{x_1, \dots, x_M\}$ , such that  $\alpha(q, x_i) \geq \alpha(q, x_{i-1})$  for all  $i \in [1, M]$ ; for notational convenience, we also define a dummy  $x_0$  with  $\alpha(q, x_0) = 0$ . Define  $y_0, y_1, \dots, y_N$  similarly for  $\beta$ , where  $N = |\mathcal{Y}|$ . Let  $t_{ij} = (\mathcal{X}_i \times \mathcal{Y}_j, z_q)$  for  $(i, j) \in [M] \times [N]$ , where  $\mathcal{X}_i = \{x_i, \dots, x_M\}$ ,  $\mathcal{Y}_j = \{y_j, \dots, y_N\}$ . Then,

$$\mathcal{T}_q = \{t_{ij} \mid (i, j) \in [M] \times [N], \alpha(q, x_i) > \alpha(q, x_{i-1}), \beta(q, y_j) > \beta(q, y_{j-1})\}.$$

Consider an arbitrary  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ . Let  $(i^*, j^*)$  be indices such that  $(x, y) = (x_{i^*}, y_{j^*})$  in the above ordering. Note that  $(x_{i^*}, y_{j^*}) \in t_{ij}$  if and only if  $1 \leq i \leq i^*$  and  $1 \leq j \leq j^*$ . Also notice that for all  $(i, j) \in [M] \times [N]$ , if  $t_{ij} \notin \mathcal{T}_q$ , then  $\sigma_{q,t_{ij}}, \tau_{q,t_{ij}} = 0$ .

$$\begin{aligned} \sum_{t \in \mathcal{T}_q: (x_{i^*}, y_{j^*}) \in t} \sigma_{q,t} \cdot \tau_{q,t} &= \sum_{i=1}^{i^*} \sum_{j=1}^{j^*} \sigma_{q,t_{ij}} \cdot \tau_{q,t_{ij}} \\ &= \sum_{i=1}^{i^*} (\alpha(q, x_i) - \alpha(q, x_{i-1})) \cdot \sum_{j=1}^{j^*} (\beta(q, y_j) - \beta(q, y_{j-1})) \\ &= \alpha(q, x_{i^*}) \cdot \beta(q, y_{j^*}) = p(q|x_{i^*}, y_{j^*}) \end{aligned}$$

as was required to prove. □

## 5 pIC Subsumes Relaxed Partition Bound

**Theorem 3.** For any relation  $f : \mathcal{X} \times \mathcal{Y} \rightarrow 2^{\mathcal{Z}}$  and constants  $\varepsilon, \delta \in [0, 1]$ ,

$$\text{pIC}(f, \varepsilon) \geq \delta \log \overline{\text{prt}}(f, \varepsilon + \delta) - (\delta \log \log(|\mathcal{X}||\mathcal{Y}|) + 3).$$

We prove this theorem in Appendix A.3. Below we summarize the main ideas.



*Proof sketch.* It is enough to show that, given a distribution  $\mathbf{p}_{XY} = \mu$  over  $\mathcal{X} \times \mathcal{Y}$ , and pseudotranscript  $Q$  such that  $\text{err}_{f,Q} \leq \varepsilon$ , there is a partition which demonstrates that  $\log \overline{\text{prt}}^\mu(f, \varepsilon + \delta) \lesssim I(X, Y; Q)/\delta$ .

The proof uses the construction from the proof of [Lemma 2](#), and modifies it carefully. Specifically, we define  $\mathbf{p}_{T|Q, X, Y}$  and  $\omega_{q,t}$  as in [Equation 8](#). Recall that we originally defined  $w$  as  $w(t) = \sum_{q \in \mathcal{Q}} \omega_{q,t}$ . Our plan now is to remove some of the weight on the tiles so that the log of the sum can be bounded by (roughly)  $I(X, Y; Q)/\delta$  as opposed to  $I_\infty(X, Y : Q)$ . Towards this, we shall define a set  $\mathcal{B}$  of “bad” pairs  $(q, t) \in \mathcal{Q} \times \mathcal{T}$  whose weights  $\omega_{q,t}$  will not be counted towards our new weight function  $w'(t)$ :

$$w'(t) = \sum_{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}} \omega_{q,t}, \quad \forall t \in \mathcal{T}.$$

The crux of the proof is to define the set  $\mathcal{B}$  such that the weight removed  $\sum_{(q,t) \in \mathcal{B}} p(q, t)$  is below  $\delta$  (it manifests as the increase in error), while keeping  $\sum_{(q,t) \notin \mathcal{B}} \omega_{q,t}$  (approximately) below  $I(X, Y; Q)/\delta$ . We show that the following choice of  $\mathcal{B}$  has both these properties:

$$\mathcal{B} = \{(q, t) \in \mathcal{Q} \times \mathcal{T} \mid \hat{\alpha}(q, t) \cdot \hat{\beta}(q, t) \geq \theta_q\}$$

where  $\hat{\alpha}(q, t) = \min_{(x,y) \in t} \alpha(q, x)$  and  $\hat{\beta}(q, t) = \min_{(x,y) \in t} \beta(q, y)$  and  $\theta_q$  is an appropriately defined threshold for each  $q \in \mathcal{Q}$  (specifically,  $\theta_q = p(q)2^\Delta$ , where  $\Delta \approx I(XY; Q)/\delta$ ).

To upper bound the mass removed, we first write  $I(XY; Q) = \sum_{q \in \mathcal{Q}, t \in \mathcal{T}} p(q, t) \varphi(q, t)$ , where  $\varphi(q, t)$  is a quantity that is lower bounded by  $\Delta$  for all  $(q, t) \in \mathcal{B}$ . This suggests the possibility of using the Markov inequality to upper bound  $\sum_{(q,t) \in \mathcal{B}} p(q, t)$ . However,  $\varphi(q, t)$  could be negative, and we cannot directly use the above expression for  $I(X, Y; Q)$  in a Markov inequality. However, we show that removing the negative terms from  $\sum_{q,t} p(q, t) \varphi(q, t)$  does not increase the sum significantly, which will let us still apply the Markov inequality.

To upper bound  $\sum_{(q,t) \notin \mathcal{B}} \omega_{q,t}$ , we use a geometric interpretation of  $\omega_{q,t}$  and the set  $\mathcal{B}$ . Fix a  $q \in \mathcal{Q}$ . Then, using the notation in the proof of [Claim 1](#), for each  $(i, j) \in [M] \times [N]$ , the tile  $t_{ij}$  will be represented by an axis-parallel rectangle on the real plane,  $R_{ij}$ , as follows.  $R_{ij}$  is defined by its diagonally opposite vertices  $(\alpha(q, x_{i-1}), \beta(q, y_{j-1}))$  and  $(\alpha(q, x_i), \beta(q, y_j))$ . (See [Figure 2](#).)  $R_{ij}$  could have zero area. These rectangles tile a rectangular region, without overlapping with each other. Further the area of the rectangle  $R_{ij}$  is the same as  $\omega_{q,t_{ij}}$ . Thus  $\sum_{t: (q,t) \notin \mathcal{B}} \omega_{q,t}$  is given by the sum of the areas of the rectangles  $R_{ij}$  for which  $(q, t_{ij}) \notin \mathcal{B}$ . The rectangles  $R_{ij}$  that correspond to  $(q, t_{ij}) \notin \mathcal{B}$  are those which have their top-right vertex (i.e.,  $(\alpha(q, x_i), \beta(q, y_j))$ ) fall “below” the hyperbola defined by the equation  $xy = \theta_q$ . Thus if  $(q, t_{ij}) \notin \mathcal{B}$ , then the entire rectangle  $R_{ij}$  is below the hyperbola  $xy = \theta_q$ . Hence the sum of their areas is upper-bounded by the area within  $R$  that is under this hyperbola, where  $R$  is the rectangle with diagonally opposite vertices  $(0, 0)$  and  $(\max_{x \in \mathcal{X}} \alpha(q, x), \max_{y \in \mathcal{Y}} \beta(q, y))$ . A calculation yields the required bound.  $\square$

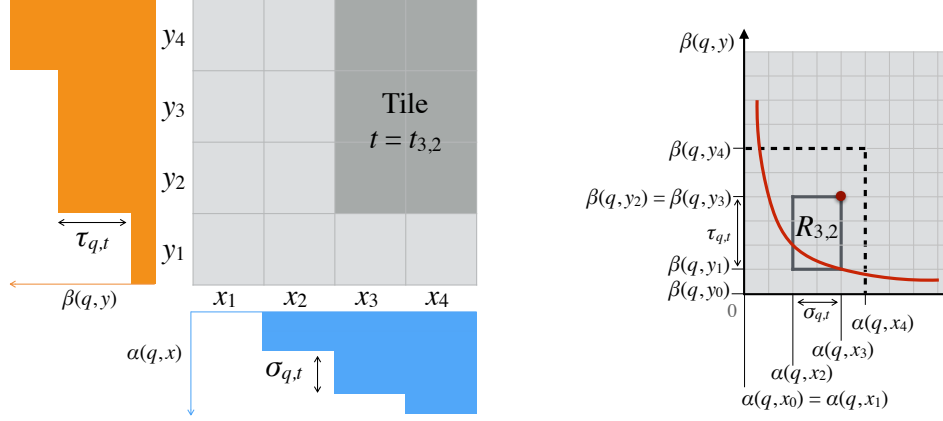
## 6 Extensions

We may define internal information complexity associated with pseudotranscripts as

$$\text{pIC}^{\text{int}}(f, \mathcal{E}) = \inf_{\substack{\text{pseudotranscript } Q: \mathbf{p}_{X,Y} \\ \text{err}_{f,Q} \leq \mathcal{E}}} \max I(X; Q|Y) + I(Y; Q|X).$$

It is easy to show that for the usual notion of information complexity (defined with respect to protocols),  $\text{IC}^{\text{int}}(f, \mathcal{E}) \leq \text{IC}(f, \mathcal{E})$ . The proof hinges on the fact that for any protocol  $\pi$  and distribution  $\mathbf{p}_{X,Y}$  on the inputs, the resulting  $\Pi$  satisfies the condition  $I(X; Y) \geq I(X; Y|\Pi)$ . However, it is unclear whether  $\text{pIC}^{\text{int}}(f, \mathcal{E})$  is necessarily upper bounded by  $\text{pIC}(f, \mathcal{E})$ . Below we define a slightly refined notion of pseudotranscripts so that information complexities defined with respect to that maintain the above inequality.

**Refined pseudotranscripts and corresponding information complexities.** A pseudotranscript  $Q$  given by  $\mathbf{p}_{Q|X,Y}$  is called a *refined pseudotranscript* if, for any distribution  $\mathbf{p}_{X,Y}$  on the inputs, it holds that  $I(X; Y) \geq I(X; Y|Q)$ . It is easy to show that for any protocol  $\pi$  and distribution  $\mathbf{p}_{X,Y}$  on the inputs, the resulting  $\Pi$  satisfies the above condition and, hence,  $\Pi$  is a refined pseudotranscript.



**Figure 2** Illustration of the geometric interpretation of  $\mathcal{B}$  used in the proof of [Theorem 3](#). The left figure shows the domain  $\mathcal{X} \times \mathcal{Y}$  and plots  $\alpha(q, x)$  and  $\beta(q, y)$  against  $x$  and  $y$ , which are sorted in the order of increasing  $\alpha(q, x)$  and  $\beta(q, y)$ , respectively (for some fixed  $q$ ). It also shows a tile  $t = t_{3,2}$  in  $\mathcal{T}_q$ , and indicates the values  $\sigma_{q,t}$  and  $\tau_{q,t}$ . The right figure shows the alternate representation of the tile  $t_{3,2}$  using the rectangular region  $R_{3,2}$ . The area of  $R_{3,2}$  equals  $\omega_{q,t_{3,2}} = \sigma_{q,t_{3,2}} \cdot \tau_{q,t_{3,2}}$ . A hyperbola corresponding to a threshold  $\theta_q$  is also shown. Since the upper-right vertex of  $R_{3,2}$ , namely the point  $(\alpha(q, x_3), \beta(q, y_2))$  is above the hyperbola,  $(q, t_{3,2}) \in \mathcal{B}$ . The area within the dotted rectangle that is under the hyperbola gives an upper-bound on the sum of areas of all rectangles under the hyperbola.

Analogous to our definition of pseudo-information complexities, we define information complexities with respect to refined pseudotranscripts

$$\begin{aligned} \hat{p}IC_{\infty}(f, \mathcal{E}) &= \inf_{\substack{\text{refined pseudotranscript } Q: \\ \text{err}_{f,Q} \leq \mathcal{E}}} I_{\infty}(X, Y : Q) \\ \hat{p}IC(f, \mathcal{E}) &= \inf_{\substack{\text{refined pseudotranscript } Q: \\ \text{err}_{f,Q} \leq \mathcal{E}}} \max_{\mathbf{p}_{X,Y}} I(X, Y; Q) \\ \hat{p}IC^{\text{int}}(f, \mathcal{E}) &= \inf_{\substack{\text{refined pseudotranscript } Q: \\ \text{err}_{f,Q} \leq \mathcal{E}}} \max_{\mathbf{p}_{X,Y}} I(X; Q|Y) + I(Y; Q|X). \end{aligned}$$

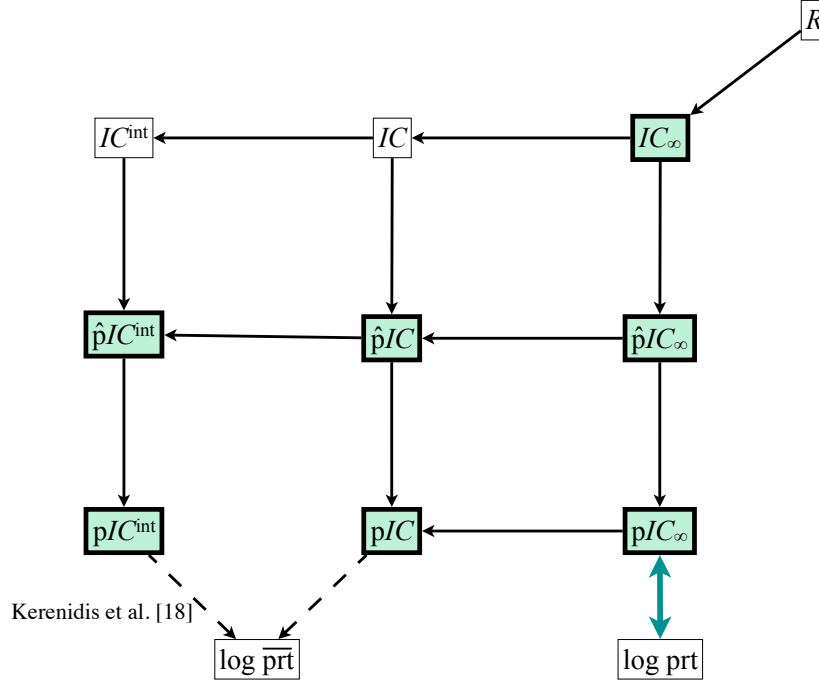
Figure 3 shows the relationship between the different complexities we consider. Since, for any protocol, its transcript is a refined pseudotranscript and refined pseudotranscripts are also pseudotranscripts, we have  $\mathbf{pX}(f, \mathcal{E}) \leq \hat{\mathbf{pX}}(f, \mathcal{E}) \leq \mathbf{X}(f, \mathcal{E})$ , where  $\mathbf{X}$  can be  $IC_{\infty}, IC$  or  $IC^{\text{int}}$ . Furthermore, analogous to  $IC^{\text{int}}(f, \mathcal{E}) \leq IC(f, \mathcal{E}) \leq IC_{\infty}(f, \mathcal{E})$ , we have  $\hat{p}IC^{\text{int}}(f, \mathcal{E}) \leq \hat{p}IC(f, \mathcal{E}) \leq \hat{p}IC_{\infty}(f, \mathcal{E})$ . Finally, in deriving a lower bound for  $IC^{\text{int}}(f, \mathcal{E})$  in terms of  $\overline{\text{prt}}(f, \mathcal{E})$  [18] only relies on the fact that the transcript (along with the public-coins)  $\Pi$  satisfies the factorization condition. Hence, the lower bound of [18] holds with  $IC^{\text{int}}$  replaced by  $\hat{p}IC^{\text{int}}$ .

## Acknowledgments

We gratefully acknowledge Mark Braverman, Prahladh Harsha, Rahul Jain, Anup Rao and the anonymous referees for helpful suggestions and pointers. The research was supported in part by NSF grant 1228856 for the first author, by ITRA, Media Lab Asia, India and by a Ramanujan fellowship from DST, India for the second author.

## References

- [1] Farid M. Abloyev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theor. Comput. Sci.*, 157(2):139–159, 1996. [3](#)
- [2] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. [1](#), [3](#), [4](#)



**Figure 3** An extended version of the map in Figure 1, including the complexity measures in Section 6.

- [3] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013. 3
- [4] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012. 2, 3
- [5] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011. 2, 3
- [6] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *APPROX-RANDOM*, pages 459–470, 2012. 3
- [7] Amit Chakrabarti, Ranganath Kondapally, and Zhenghui Wang. Information complexity versus corruption and applications to orthogonality and gap-hamming. In *APPROX-RANDOM*, pages 483–494, 2012. 3
- [8] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278, 2001. 1, 3
- [9] Lila Fontes, Rahul Jain, Iordanis Kerenidis, Sophie Laplante, Mathieu Lauri  re, and J  r  mie Roland. Relative discrepancy does not separate information and communication complexity. pages 506–516, 2015. 3
- [10] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *FOCS*, pages 176–185, 2014. 2, 3
- [11] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of communication and external information. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2015. 2, 3
- [12] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010. 3
- [13] Siu-Wai Ho and Sergio Verd  . Convexity/concavity of R  nyi entropy and  $\alpha$ -mutual information. In *Information Theory (ISIT), 2015 IEEE International Symposium on*, pages 745–749, 2015. 4, 6

- [14] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *IEEE Conference on Computational Complexity*, pages 247–258, 2010. [1](#), [3](#), [5](#)
- [15] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003. [3](#)
- [16] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *IEEE Conference on Computational Complexity*, pages 285–296, 2005. [3](#)
- [17] T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *STOC*, pages 673–682, 2003. [3](#)
- [18] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *FOCS*, pages 500–509, 2012. [2](#), [3](#), [5](#), [10](#), [14](#)
- [19] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, New York, 1997. [3](#)
- [20] Stephen J Ponzio, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62(2):323–355, 2001. [3](#)
- [21] Manoj Prabhakaran and Vinod M. Prabhakaran. Tension bounds for information complexity. *CoRR*, abs/1408.6285, 2014. [4](#)
- [22] Alfred Rényi. On measures of information and entropy. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561, 1960. [3](#)
- [23] Michael E. Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *STOC*, pages 360–369, 2002. [3](#)
- [24] R. Sibson. Information radius. *Z. Wahrscheinlichkeitstheorie und Verw. Geb.*, 14:149–161, 1969. [4](#)
- [25] Sergio Verdú.  $\alpha$ -mutual information. In *Information Theory and Applications Workshop (ITA)*, 2015. [2](#), [4](#), [5](#)
- [26] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *STOC*, pages 209–213, 1979. [1](#), [3](#)
- [27] Moshe Zakai and Jacob Ziv. A generalization of the rate-distortion theory and applications. In *Information Theory New Trends and Open Problems*, pages 87–123. Springer, 1975. [4](#)
- [28] Jacob Ziv and Moshe Zakai. On functionals satisfying a data-processing theorem. *Information Theory, IEEE Transactions on*, 19(3):275–283, 1973. [4](#)

## A Omitted Proofs

### A.1 $I(A; B) \leq I_\infty(A; B)$

For the sake of completeness, we include a proof that  $I(A; B) \leq I_\infty(A; B)$ .

$$\begin{aligned}
I_\infty(A; B) &= \log \left( \sum_{b \in \mathcal{B}} \max_{a \in \mathcal{A}: \mathbf{p}_A(a) > 0} \mathbf{p}_{B|A}(b|a) \right) \\
&\geq \log \left( \sum_{b \in \mathcal{B}: \mathbf{p}_B(b) > 0} \mathbf{p}_B(b) \max_{a \in \mathcal{A}: \mathbf{p}_A(a) > 0} \frac{\mathbf{p}_{B|A}(b|a)}{\mathbf{p}_B(b)} \right) \\
&\geq \log \left( \sum_{b \in \mathcal{B}: \mathbf{p}_B(b) > 0} \mathbf{p}_B(b) \sum_{a \in \mathcal{A}: \mathbf{p}_A(a) > 0} \mathbf{p}_{A|B}(a|b) \frac{\mathbf{p}_{B|A}(b|a)}{\mathbf{p}_B(b)} \right) \\
&= \log \left( \sum_{a \in \mathcal{A}, b \in \mathcal{B}: \mathbf{p}_A(a) > 0, \mathbf{p}_B(b) > 0} \mathbf{p}_{A,B}(a, b) \frac{\mathbf{p}_{B|A}(b|a)}{\mathbf{p}_B(b)} \right) \\
&\geq \sum_{a \in \mathcal{A}, b \in \mathcal{B}: \mathbf{p}_A(a) > 0, \mathbf{p}_B(b) > 0} \mathbf{p}_{A,B}(a, b) \log \left( \frac{\mathbf{p}_{B|A}(b|a)}{\mathbf{p}_B(b)} \right) \\
&= I(A; B).
\end{aligned}$$

### A.2 Proof of Lemma 1

*Proof.* Consider the weight function  $w : \mathcal{T} \rightarrow [0, 1]$  that satisfies the conditions (1) and (2) such that  $\text{prt}(f, \mathcal{E}) = \sum_{t \in \mathcal{T}} w(t)$ . Define the random variable  $Q$  over  $\mathcal{Q} = \mathcal{T}$  such that  $\mathbf{p}_{Q|XY}(t|x, y) = w(t)$  if  $(x, y) \in t$  and 0 otherwise. Note that this is a valid probability distribution since for all  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ , we have

$$\sum_{t \in \mathcal{Q}} \mathbf{p}_{Q|XY}(t|x, y) = \sum_{t \in \mathcal{Q}: (x, y) \in t} w(t) = 1.$$

Let  $a_t, b_t \geq 0$  be such that  $a_t \cdot b_t = w(t)$  (for instance,  $a_t = b_t = \sqrt{w(t)}$ ), and define functions  $\alpha : \mathcal{Q} \times \mathcal{X} \rightarrow \mathbb{R}^+$  and  $\beta : \mathcal{Q} \times \mathcal{Y} \rightarrow \mathbb{R}^+$  as follows:

$$\alpha(t, x) = \begin{cases} a_t & \text{if } x \in \mathcal{X}_t \\ 0 & \text{otherwise} \end{cases} \quad \beta(t, y) = \begin{cases} b_t & \text{if } y \in \mathcal{Y}_t \\ 0 & \text{otherwise} \end{cases}$$

Then,  $\mathbf{p}_{Q|XY}(t|x, y) = \alpha(t, x) \cdot \beta(t, y)$ , and hence it satisfies the factorization condition. Further, for each  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ ,

$$\text{err}_{f,Q}(x, y) = \sum_{t \in \mathcal{Q}: z_t \notin f(x, y)} \mathbf{p}_{Q|XY}(t|x, y) = \sum_{t \in \mathcal{Q}: (x, y) \in t, z_t \notin f(x, y)} w(t) \leq \mathcal{E}(x, y).$$

Hence  $\text{pIC}_\infty(f, \mathcal{E}) \leq I_\infty(X, Y; Q)$ . On the other hand,

$$I_\infty(X, Y; Q) = \log \sum_{t \in \mathcal{Q}} \max_{x, y} \mathbf{p}_{Q|XY}(t|x, y) = \log \sum_{t \in \mathcal{T}} w(t) = \log \text{prt}(f, \mathcal{E}),$$

concluding the proof.  $\square$

### A.3 Proof of Theorem 3

*Proof.* We shall show that for any distribution  $\mathbf{p}_{XY} = \mu$  over  $\mathcal{X} \times \mathcal{Y}$ , and any pseudotranscript  $Q$  such that  $\text{err}_{f,Q} \leq \varepsilon$  (i.e.,  $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$ ,  $\text{err}_{f,Q}(x, y) \leq \varepsilon$ ),  $I(X, Y; Q) \geq \delta \log \text{prt}^\mu(f, \varepsilon + \delta) - (\delta \log \log |\mathcal{X}||\mathcal{Y}| + 3)$ . This gives the desired result, since

$$\text{pIC}(f, \varepsilon) = \inf_{Q: \text{err}_{f,Q} \leq \varepsilon} \max_{\mathbf{p}_{XY}} I(X, Y; Q) \geq \max_{\mathbf{p}_{XY}} \inf_{Q: \text{err}_{f,Q} \leq \varepsilon} I(X, Y; Q)$$

and as shown in [18],  $\overline{\text{prt}}(f, \varepsilon') = \max_{\mu} \overline{\text{prt}}^{\mu}(f, \varepsilon')$ .

The proof uses the construction from the proof of Lemma 2, and modifies it carefully. Specifically, we define  $\mathbf{p}_{T|Q,X,Y}$  and  $\omega_{q,t}$  as before. (Note that since we are now given a distribution  $\mu$  for the random variables  $(X, Y)$ , this also gives us a full distribution  $\mathbf{p}_{Q,T,X,Y}$ ; below  $p(x, y) = \mu(x, y)$ .) Recall that we originally defined  $w$  as  $w(t) = \sum_{q \in \mathcal{Q}} \omega_{q,t}$ . Our plan now is to remove some of the weight on the tiles so that the log of the sum can be bounded by (roughly)  $I(X, Y; Q)/\delta$  as opposed to  $I_{\infty}(X, Y : Q)$ . Towards this, we shall define a set  $\mathcal{B}$  of “bad” pairs  $(q, t) \in \mathcal{Q} \times \mathcal{T}$  whose weights  $\omega_{q,t}$  will not be counted towards  $w'(t)$ :

$$w'(t) = \sum_{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}} \omega_{q,t}, \quad \forall t \in \mathcal{T}.$$

While defining  $\mathcal{B}$ , we need to ensure that the weight removed increases the *average* error  $\overline{\text{err}}_{f,w'}^{\mu}$  by at most  $\delta$  compared to  $\overline{\text{err}}_{f,w}^{\mu} = \overline{\text{err}}_{f,Q}^{\mu} = \varepsilon$ .

We define parameters  $\Delta = (I(XY; Q) + 1)/\delta$  and for each  $q \in \mathcal{Q}$ ,  $\theta_q = p(q)2^{\Delta}$ . Let  $\hat{\alpha}(q, t) = \min_{(x,y) \in t} \alpha(q, x)$  and  $\hat{\beta}(q, t) = \min_{(x,y) \in t} \beta(q, y)$ . Then we define

$$\mathcal{B} = \{(q, t) \in \mathcal{Q} \times \mathcal{T} \mid \hat{\alpha}(q, t) \cdot \hat{\beta}(q, t) \geq \theta_q\}$$

We make the following claims, which we prove in Appendix A.3.1 and Appendix A.3.2.

**Claim 2.**  $\sum_{(q,t) \in \mathcal{B}} p(q, t) \leq \delta$ .

**Claim 3.**  $\log \sum_{(q,t) \notin \mathcal{B}} \omega_{q,t} \leq \Delta + \log \log(|\mathcal{X}||\mathcal{Y}|) + 2$ .

Using these claims, we complete the proof. Firstly, note that  $w'(t) \leq w(t)$  for every  $t \in \mathcal{T}$  and, since  $w$  satisfies condition (1),  $w'$  satisfies condition (3). Also, from Claim 2 it follows that

$$\begin{aligned} \overline{\text{err}}_{f,w'}^{\mu} &= 1 - \sum_{x,y} p(x, y) \sum_{\substack{t \in \mathcal{T}: (x,y) \in t, \\ z_t \in f(x,y)}} w'(t) = 1 - \sum_{x,y} p(x, y) \sum_{\substack{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}: \\ (x,y) \in t, \\ z_t \in f(x,y)}} \omega_{q,t} \\ &= 1 - \sum_{x,y} p(x, y) \sum_{\substack{(q,t) \in \mathcal{Q} \times \mathcal{T}: \\ (x,y) \in t, \\ z_t \in f(x,y)}} \omega_{q,t} + \sum_{x,y} p(x, y) \sum_{\substack{(q,t) \in \mathcal{B}: \\ (x,y) \in t, \\ z_t \in f(x,y)}} \omega_{q,t} \\ &= \overline{\text{err}}_{f,w}^{\mu} + \sum_{(q,t) \in \mathcal{B}} \sum_{\substack{(x,y) \in t, \\ z_t \in f(x,y)}} p(x, y) \omega_{q,t} \leq \overline{\text{err}}_{f,w}^{\mu} + \sum_{(q,t) \in \mathcal{B}} \sum_{(x,y) \in t} p(x, y) \omega_{q,t} \\ &= \overline{\text{err}}_{f,w}^{\mu} + \sum_{(q,t) \in \mathcal{B}} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) p(q, t | x, y) \quad \text{by (6)} \\ &= \overline{\text{err}}_{f,w}^{\mu} + \sum_{(q,t) \in \mathcal{B}} p(q, t) \leq \varepsilon + \delta \quad \text{by Claim 2} \end{aligned}$$

Hence,

$$\begin{aligned} \log \overline{\text{prt}}^{\mu}(f, \varepsilon + \delta) &\leq \sum_{t \in \mathcal{T}} w'(t) = \log \sum_{(q,t) \notin \mathcal{B}} \omega_{q,t} \\ &\leq \Delta + \log \log |\mathcal{X}||\mathcal{Y}| + 2 \quad \text{by Claim 3} \\ &= \frac{I(X, Y; Q)}{\delta} + \frac{1}{\delta} + \log \log |\mathcal{X}||\mathcal{Y}| + 2 \\ &\leq \frac{I(X, Y; Q)}{\delta} + \log \log |\mathcal{X}||\mathcal{Y}| + \frac{3}{\delta} \quad \text{since } \delta \in [0, 1] \end{aligned}$$

That is,  $I(X, Y; Q) \geq \delta \log \overline{\text{prt}}^{\mu}(f, \varepsilon + \delta) + (\delta \log \log |\mathcal{X}||\mathcal{Y}| + 3)$ , as was required to prove.  $\square$

The proofs of the two claims used above follow.



### A.3.1 Proof of Claim 2

*Proof.* This claim follows from Markov's inequality applied to an appropriate random variable, whose mean is related to  $I(XY; Q)$ . First, we expand  $I(XY; Q)$  as follows:

$$\begin{aligned}
I(XY; Q) &= \sum_{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}} p(q, x, y) \log \frac{p(q|x, y)}{p(q)} \\
&= \sum_{q \in \mathcal{Q}, t \in \mathcal{T}, x \in \mathcal{X}, y \in \mathcal{Y}} p(q, t, x, y) \log \frac{p(q|x, y)}{p(q)} \\
&= \sum_{q \in \mathcal{Q}, t \in \mathcal{T}} p(q, t) \sum_{(x, y) \in t} p(x, y|q, t) \log \frac{p(q|x, y)}{p(q)} \quad \text{since } (x, y) \notin t \implies p(q, t, x, y) = 0 \\
&= \sum_{q \in \mathcal{Q}, t \in \mathcal{T}} p(q, t) \varphi(q, t)
\end{aligned}$$

where we have defined

$$\varphi(q, t) = \begin{cases} \sum_{(x, y) \in t} p(x, y|q, t) \log \frac{p(q|x, y)}{p(q)} & \text{if } p(q, t) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

That is,  $\varphi(q, t)$  is the average value of  $\log \frac{p(q|x, y)}{p(q)}$  averaged over all  $(x, y) \in t$  using the distribution  $\mathbf{p}_{XY|Q=q, T=t}$ . We note that for all  $(q, t) \in \mathcal{B}$ ,  $\varphi(q, t) \geq \Delta$ , since for each  $(x, y) \in t$ ,  $p(q|x, y) = \alpha(q, x)\beta(q, y) \geq \hat{\alpha}(q, t)\hat{\beta}(q, t) \geq \theta_q$  and hence  $\log \frac{p(q|x, y)}{p(q)} \geq \log \frac{\theta_q}{p(q)} = \Delta$ . This suggests the possibility of using the Markov inequality to bound  $\sum_{(q, t) \in \mathcal{B}} p(q, t)$ . However,  $\varphi(q, t)$  could be negative, and we cannot directly use the above expression for  $I(X, Y; Q)$  in a Markov inequality. However, we claim that removing the negative terms from  $\sum_{q, t} p(q, t) \varphi(q, t)$  does not increase the sum significantly, which will let us still apply the Markov inequality.

More precisely, let  $\mathcal{D} = \{(q, t) \in \mathcal{Q} \times \mathcal{T} \mid \min_{(x, y) \in t} p(q|x, y) \geq p(q)\}$ . Note that if  $(q, t) \in \mathcal{D}$ , then  $\varphi(q, t) \geq 0$ . We claim that

$$I(X, Y; Q) \geq \left( \sum_{(q, t) \in \mathcal{D}} p(q, t) \varphi(q, t) \right) - 1. \quad (9)$$

Assuming (9), we can conclude the proof of the claim as follows. Note that  $\mathcal{B} \subseteq \mathcal{D}$  since if  $(q, t) \in \mathcal{B}$ ,  $\min_{(x, y) \in t} p(q|x, y) = \hat{\alpha}(q, t) \cdot \hat{\beta}(q, t) \geq \theta_q \geq p(q)$ . Also, recall that for  $(q, t) \in \mathcal{B}$ ,  $\varphi(q, t) \geq \Delta$ . Hence,

$$\delta \Delta = I(X, Y; Q) + 1 \geq \sum_{(q, t) \in \mathcal{D}} p(q, t) \varphi(q, t) \geq \Delta \sum_{(q, t) \in \mathcal{B}} p(q, t),$$

and therefore  $\sum_{(q, t) \in \mathcal{B}} p(q, t) \leq \delta$ .

To prove (9), consider again the expansion of  $I(X, Y; Q)$  as

$$I(XY; Q) = \left( \sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) \geq p(q)}} p(q, x, y) \log \frac{p(q|x, y)}{p(q)} \right) - \left( \sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) < p(q)}} p(q, x, y) \log \frac{p(q)}{p(q|x, y)} \right).$$

in which all the terms within each summation is non-negative. To bound the second term, writing  $\eta = \sum_{q, x, y: p(q|x, y) < p(q)} p(q, x, y)$ , we use Jensen's inequality to write

$$\begin{aligned}
\sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) < p(q)}} p(q, x, y) \log \frac{p(q)}{p(q|x, y)} &\leq \eta \log \sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) < p(q)}} \frac{p(q, x, y)}{\eta} \cdot \frac{p(q)}{p(q|x, y)} \\
&= \eta \log \frac{1}{\eta} + \eta \log \sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) < p(q)}} p(x, y) p(q) \\
&\leq \eta \log \frac{1}{\eta} \leq \frac{\log e}{e} < 1.
\end{aligned}$$

where to get to the last line we used the fact that  $\sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) < p(q)}} p(x, y)p(q) \leq \sum_{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y)p(q) = 1$ . Hence

$$\begin{aligned} I(XY; Q) &\geq \left( \sum_{\substack{q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}: \\ p(q|x, y) \geq p(q)}} p(q, x, y) \log \frac{p(q|x, y)}{p(q)} \right) - 1 \\ &\geq \left( \sum_{(q, t) \in \mathcal{D}, (x, y) \in t} p(q, t, x, y) \log \frac{p(q|x, y)}{p(q)} \right) - 1 \quad \text{since } (x, y) \in t, (q, t) \in \mathcal{D} \implies p(q|x, y) \geq p(q) \\ &= \left( \sum_{(q, t) \in \mathcal{D}} p(q, t) \varphi(q, t) \right) - 1 \end{aligned}$$

completing the proof of (9) and of the claim.  $\square$

### A.3.2 Proof of Claim 3

*Proof.* We need to upper-bound

$$\sum_{\substack{q \in \mathcal{Q}, t \in \mathcal{T}: \\ (q, t) \notin \mathcal{B}}} \omega_{q, t} = \sum_{q \in \mathcal{Q}} \sum_{\substack{t \in \mathcal{T}_q: \\ (q, t) \notin \mathcal{B}}} \sigma_{q, t} \tau_{q, t}.$$

For this we shall use a geometric interpretation of this sum.

Fix  $q \in \mathcal{Q}$ . Recall from the proof of Claim 1, that for each  $q$ , we order  $\mathcal{X} = \{x_1, \dots, x_M\}$  and  $\mathcal{Y} = \{y_1, \dots, y_N\}$  such that  $\alpha(q, x_i) \geq \alpha(q, x_{i-1})$  and  $\beta(q, y_j) \geq \beta(q, y_{j-1})$  (taking  $\alpha(q, x_0) = \beta(q, y_0) = 0$ ), and  $t_{ij} = (\mathcal{X}_i \times \mathcal{Y}_j, z_q)$  for  $(i, j) \in [M] \times [N]$ , where  $\mathcal{X}_i = \{x_i, \dots, x_M\}$ ,  $\mathcal{Y}_j = \{y_j, \dots, y_N\}$ . Then

$$\mathcal{T}_q = \{t_{ij} \mid (i, j) \in [M] \times [N], \alpha(q, x_i) > \alpha(q, x_{i-1}), \beta(q, y_j) > \beta(q, y_{j-1})\}.$$

Consider the rectangular region  $R \subseteq \mathbb{R}^2$  defined by the diagonally opposite vertices  $(0, 0)$  and  $(\alpha_q^*, \beta_q^*)$ , where  $\alpha_q^* = \max_{x \in \mathcal{X}} \alpha(q, x)$  and  $\beta_q^* = \max_{y \in \mathcal{Y}} \beta(q, y)$ . For each  $(i, j) \in [M] \times [N]$  let the (possibly empty) rectangular region  $R_{ij}$  be defined by opposite vertices  $(\alpha(q, x_{i-1}), \beta(q, y_{j-1}))$  and  $(\alpha(q, x_i), \beta(q, y_j))$ . (See Figure 2.) Then note that the entire region  $R$  is tiled by the rectangles  $R_{ij}$ , without any overlap:

$$R = \bigcup_{(i, j) \in [M] \times [N]} R_{ij} \quad (i, j) \neq (i', j') \implies R_{ij} \cap R_{i'j'} = \emptyset.$$

Further, the area of the rectangle  $R_{ij}$  is the same as  $\omega_{q, t_{ij}} = \sigma_{q, t_{ij}} \tau_{q, t_{ij}} = (\alpha(q, x_i) - \alpha(q, x_{i-1})) (\beta(q, y_j) - \beta(q, y_{j-1}))$ . Thus,

$$\sum_{\substack{t \in \mathcal{T}_q: \\ (q, t) \notin \mathcal{B}}} \omega_{q, t} = \sum_{\substack{(i, j) \in [M] \times [N]: \\ (q, t_{ij}) \notin \mathcal{B}}} \text{area}(R_{ij}).$$

Now we need to identify the rectangles  $R_{ij}$  such that  $(q, t_{ij}) \notin \mathcal{B}$ . Firstly, recall that  $\hat{\alpha}(q, t_{ij}) = \min_{(x, y) \in t_{ij}} \alpha(q, x) = \alpha(q, x_i)$ , and similarly  $\hat{\beta}(q, t_{ij}) = \beta(q, y_j)$ . Hence  $(q, t_{ij}) \in \mathcal{B}$  if and only if  $\alpha(q, x_i) \beta(q, y_j) \geq \theta_q$ . In terms of the rectangle  $R_{ij}$  this corresponds to having its top-right vertex (i.e.,  $(\alpha(q, x_i), \beta(q, y_j))$ ) fall “above” the hyperbola defined by the equation  $xy = \theta_q$ . Thus if  $(q, t_{ij}) \notin \mathcal{B}$ , then the entire rectangle  $R_{ij}$  is below the hyperbola  $xy = \theta_q$ . The sum of their areas is upper-bounded by the area within  $R$  that is under this hyperbola.

We consider two cases for  $q$ : when the hyperbola intersects  $R$  and when it does not; the latter happens when  $\theta_q > \alpha_q^* \beta_q^*$ . Let  $\mathcal{S} = \{q \mid \theta_q > \alpha_q^* \beta_q^*\}$ . If  $q \in \mathcal{S}$ , then clearly the area of  $R$  below the hyperbola is the entire area,  $\alpha_q^* \beta_q^*$ . Otherwise, the area under the hyperbola is found by integration as

$$\theta_q + \int_{\frac{\theta_q}{\beta_q^*}}^{\alpha_q^*} \frac{\theta_q}{x} dx = \theta_q + \theta_q \ln \frac{\alpha_q^* \beta_q^*}{\theta_q},$$

where  $\ln$  stands for natural logarithm.

Let  $\lambda = \sum_{q \notin \mathcal{S}} p(q)$ . Then,

$$\begin{aligned}
\sum_{\substack{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}: \\ q \in \mathcal{S}}} \omega_{q,t} &= \sum_{q \in \mathcal{S}} \alpha_q^* \beta_q^* \leq \sum_{q \in \mathcal{S}} \theta_q = (1 - \lambda) 2^\Delta \\
\sum_{\substack{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}: \\ q \notin \mathcal{S}}} \omega_{q,t} &\leq \sum_{q \in \mathcal{Q} \setminus \mathcal{S}} \theta_q + \theta_q \ln \frac{\alpha_q^* \beta_q^*}{\theta_q} \\
&= \lambda 2^\Delta + \lambda 2^\Delta \sum_{q \in \mathcal{Q} \setminus \mathcal{S}} \frac{p(q)}{\lambda} \ln \frac{\alpha_q^* \beta_q^*}{p(q) 2^\Delta} \\
&\leq \lambda 2^\Delta + \lambda 2^\Delta \ln \sum_{q \in \mathcal{Q} \setminus \mathcal{S}} \frac{\alpha_q^* \beta_q^*}{\lambda 2^\Delta} && \text{By Jensen's inequality} \\
&\leq \lambda 2^\Delta + \lambda 2^\Delta \ln \left( \sum_{q \in \mathcal{Q}} \alpha_q^* \beta_q^* \right) + \lambda 2^\Delta \ln \frac{1}{\lambda 2^\Delta} \\
&\leq \lambda 2^\Delta + 2^\Delta \cdot I_\infty(X, Y : Q) \cdot \ln 2 + \frac{1}{e} && \text{since for all } a > 0, a \ln \frac{1}{a} \leq \frac{1}{e} \\
&\leq \lambda 2^\Delta + 2^\Delta \cdot \log |\mathcal{X}| |\mathcal{Y}| \cdot \ln 2 + \frac{1}{e} && \text{since } I_\infty(X, Y : Q) \leq \log |\mathcal{X}| |\mathcal{Y}| \\
\sum_{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}} \omega_{q,t} &\leq 2^\Delta (1 + \log |\mathcal{X}| |\mathcal{Y}| \cdot \ln 2 + \frac{1}{e}) \\
&\leq 2^\Delta (4 \log |\mathcal{X}| |\mathcal{Y}|) && \text{since } |\mathcal{X}| |\mathcal{Y}| \geq 2
\end{aligned}$$

Note that we assumed  $|\mathcal{X}| |\mathcal{Y}| \geq 2$ , because otherwise  $|\mathcal{X}| = |\mathcal{Y}| = 1$  and the theorem holds trivially (with LHS being 0 and RHS being negative). From the above we obtain that  $\log \sum_{(q,t) \in (\mathcal{Q} \times \mathcal{T}) \setminus \mathcal{B}} \omega_{q,t} \leq \Delta + \log \log |\mathcal{X}| |\mathcal{Y}| + 2$  completing the proof of the claim.  $\square$